THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 3, March 1979
Printed in U.S.A.

Objective and Subjective Performance of Tandem Connections of Waveform Coders with an LPC Vocoder

By D. J. GOODMAN, C. SCAGLIOLA, R. E. CROCHIERE, L. R. RABINER, and J. GOODMAN

(Manuscript received June 9, 1978)

In a recently proposed communication system, there would be tandem connections of 16 kb/s delta modulators and 2.4 kb/s vocoders. Preliminary work has indicated that such tandem links would be of substantially lower quality than either the delta modulator link or the vocoder link alone. The present study, which includes an elaborate subjective speech quality experiment, confirms this preliminary conclusion. It also shows that two other differential waveform coders are no better than the proposed delta modulator in tandem links. On the other hand, a 5-band sub-band coder does offer substantially higher quality than the delta modulator. Still, its performance in tandem with the vocoder is poorer than that of the vocoder or the sub-band coder alone and is probably of only marginal value for practical communication. We have obtained several objective measures of speech quality which, for the most part, show relatively little correlation with subjective quality. The most successful objective predictor of subjective ratings is a linear combination of linear predictive coding distances.

I. BACKGROUND AND INTRODUCTION

1.1 Background

Recent plans for United States government digital communication networks have focused attention on the compatibility of 2.4-kb/s (narrowband) vocoder systems and 16-kb/s (wideband) waveform coding schemes. An important question arises in the implementation of such a system: If both narrowband and wideband systems are designed to provide adequate speech communication individually, will a tandem connection of them also function adequately?

A recent study of this question,[1, 2] using signal-to-noise ratio (s/n) and a spectral distance measure as criteria of merit, has cast doubt on the viability of circuits containing a 2.4-kb/s LPC (linear predictive coding) vocoder and a 16-kb/s CVSD (continuously variable slope delta modulation) waveform coder. In that study, it appeared that CVSD was the weak link in these tandem connections. However, the conclusions could only be regarded as tentative because the speech material included in the study was very limited and because the relationship of the objective performance measures to the quality of communication experienced by human users was by no means evident.

1.2 Aims

In the work reported here, we extend previous results by:

(*i*) Studying three 16-kb/s waveform coders in addition to CVSD.

(*ii*) Presenting subjective as well as objective performance measures.

(*iii*) Greatly enlarging the variety of speech material processed by the various communication systems.

The specific questions addressed in our study are:

(*i*) What is the subjective quality of tandem connections of narrowband and wideband systems, relative to the quality of individual systems?

(*ii*) Are there alternatives to CVSD that offer higher quality in either (or both) individual or tandem performance?

(*iii*) What is the relationship of objective measures of system performance to subjective assessment of speech quality?

1.3 An experiment

To answer these questions we produced, in software on a Data General Eclipse computer, a 2.4-kb/s LPC vocoder and four different 16-kb/s waveform encoders. They are:

(*i*) The CVSD studied in Refs. 1 and 2.

(*ii*) A double integration version of CVSD, which we call ADM (adaptive delta modulator).

(*iii*) A two-bit 8 kHz ADPCM (adaptive differential PCM).

(*iv*) SBC (sub-band coding) with five separate channels spanning the 200 to 3200-Hz band of speech energy.

Relative to CVSD, ADM has virtually the same circuit complexity (requiring only one additional resistor and one capacitor), ADPCM is perhaps 2 to 3 times as complex, and SBC is approximately 10 to 20 times as complicated.

The five coding schemes (four waveform coders and LPC) were used in 13 different communication systems (five coders individually, four waveform coders preceding LPC, four waveform coders following LPC). These systems processed a total of 148 speech samples from four

talkers (two male and two female) at three different power levels (spanning a 30-dB range).

Twenty-two subjects rated each of the processed speech samples on a 9-point scale. Each sample consisted of one sentence of 2 to 3 seconds duration, and no sentence was heard more than once by any individual subject. Although the subjects were asked to rate overall speech quality, it is felt that intelligibility had a greater influence over their ratings than it does in experiments in which a few sentences are repeated many times.

In addition, four different objective measures of system quality were calculated. These include the s/n and spectral distance measure used in Refs. 1 and 2, and also two segmental signal-to-noise ratios[3] that have been shown in other work to be more closely related to subjective quality than s/n.[4]

Statistical analysis of the subjective data reveal a complicated pattern of interactions among the experimental variables. The relative performances of the various coding schemes are dependent in a complicated way on talker and on input level as well as on (individual or tandem) system configuration. In spite of the complicated dependence of subjective quality on physical conditions, clear patterns in the data emerge to answer our original questions.

Among the individual circuits, SBC has on the average the highest subjective quality, followed by LPC, CVSD, ADM, and ADPCM, in that order. Tandem connections all are substantially degraded relative to individual circuits. Among the waveform coders, SBC provides the best tandem connections with LPC, but the SBC-LPC tandems are substantially worse than either individual system. The tandems involving the other waveform coders are probably inadequate for effective speech communication.

Among the objective measures, s/n as in other studies[3,4] was found to be very poorly correlated with subjective quality. Moreover, with the diversity of circuit conditions and speech material presented here, the segmental signal-to-noise ratios were also of little use in predicting subjective quality. The spectral distance measure was the only one that was somewhat useful: a linear regression model based on distance measures of both tandem links and on overall distance accounted for 60 percent of the variance in the average ratings.

II. SYSTEM DESCRIPTION

2.1 Overview

In the narrowband-to-wideband tandem link shown in Fig. 1, the input speech appears as 16-bit PCM with 8-kHz sampling rate. It is first bandpass filtered to a bandwidth of 200 to 3200 Hz by means of a 6th order elliptic bandpass filter. It is then vocoded by the LPC vocoder. At

the output of the vocoder, the sampling rate is converted (if necessary) by digital techniques[5] to the sampling rate of the waveform coder. This conversion has no effect on the tandem connection and is virtually "transparent" in terms of quality. The gains G and $1/G$ before and after the coder are used in measuring the dynamic range (i.e., variations in performance as a function of signal level) of the waveform coder. The output of the coder is lowpass filtered to 3200 Hz, and its sampling rate is converted back to 8 kHz and the output signal is processed by a 3200-Hz lowpass filter. In Fig. 2, the same signal processing operations are shown with the ordering that provides a wideband-to-narrowband connection.

2.2 The narrowband system (LPC)

The narrowband system consists of a linear predictive coding (LPC) system based on an all-pole model of the speech production mechanism. The all-pole model implies that, within a frame of speech, the output speech sequence is approximated by

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + G' u_n, \tag{1}$$

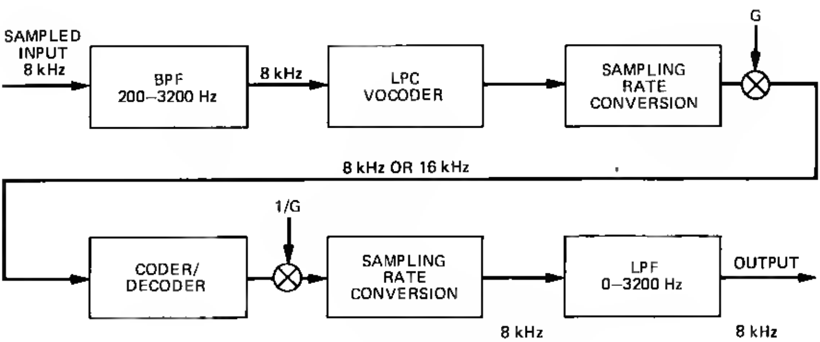

Fig. 1—Narrowband-to-wideband system.

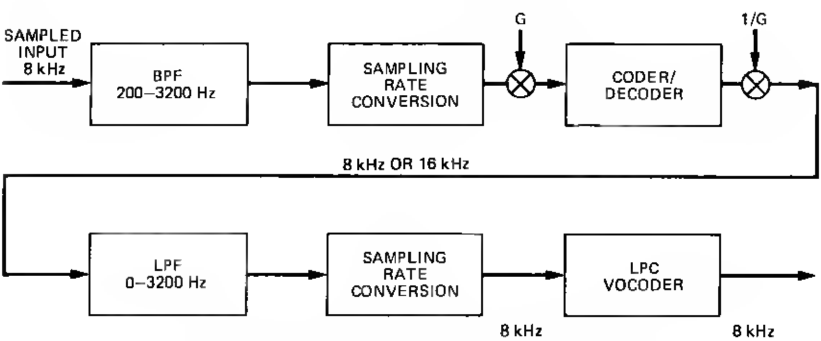

Fig. 2—Wideband-to-narrowband system.

where p is the number of poles, u_n is the appropriate input, G' is the gain, and the a_k's are the LPC coefficients that represent the spectral characteristics of the speech frame. For a voiced speech segment, u_n is a sequence of pulses separated by the pitch period. If the segment is unvoiced, pseudorandom white noise is used as input.

In our study, the LPC coefficients were calculated by the autocorrelation method with $p = 12$. The analysis was performed every 20 ms (50 times/s) with a variable analysis frame size. The frame size was proportional to a running average of the pitch period as obtained at the pitch detector output.[6] A Hamming window was used prior to the LPC analysis. Pitch detection and voiced-unvoiced (V/U) analysis were done using the modified autocorrelation method.

For quantization purposes, the LPC coefficients were converted to log area ratio coefficients, which were coded by means of ADPCM techniques.[8] An overall bit rate of 2.4 kb/s was obtained by allocating 48 bits to each of the 50 frames per second. Details of the encoding scheme are given in the references.

2.3 Delta modulators, CVSD and ADM

The experiment includes two delta modulators, CVSD, and a double integration version of CVSD which we refer to as ADM. Both of them can be represented by the block diagram in Fig. 3. Their principal difference is in the nature of the signal feedback path which is a single integrator in CVSD and a double integrator in ADM.

In both coders, the step size voltage can be generated by an RC integrator as described in Ref. 1. The integrator input depends on the three most recent output bits. If they are identical, the step size increases; otherwise, it decreases. The adaptation equation is

$$\Delta_{k+1} = \beta\Delta_k + (1 - \beta)(V_k + \Delta_{\min}), \tag{2}$$

where

Δ_k is the step size at the kth sampling instant,

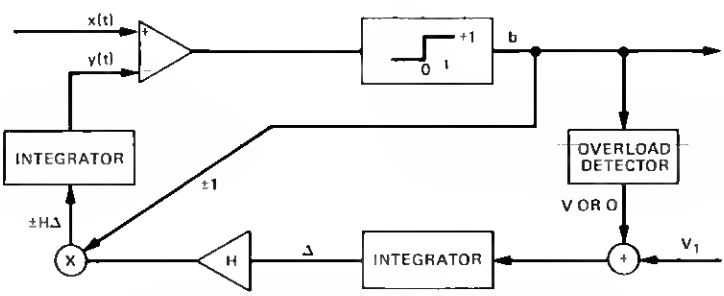

Fig. 3—Block diagram of the CVSD and ADM coders.

$\beta = 0.99$ is the step size leakage constant corresponding to an RC time constant of 6.4 ms,
Δ_{min} is the minimum step size, and
$V_k = \Delta_{max}$ wben the three most recent outputs are identical and
$V_k = 0$ otherwise.

The dynamic range of the coder is determined by $\Delta_{max}/\Delta_{min}$, which is 150 (44 dB) in the CVSD and 256 (48 dB) in the ADM.

2.3.1 CVSD

As in Ref. 1, the signal feedback loop is a single integrator with a 1-ms time constant. The difference equation is

$$y_{k+1} = \alpha_1 y_k + H(1-\alpha_1)b_k\Delta_k, \tag{3}$$

where
y_k is the integrator output at the kth sampling instant,
$\alpha_1 = 0.94$ is the integrator leakage constant, corresponding to an RC time constant of 1 ms,
$H = 3$ is the integrator gain, and
$b_k = \pm 1$ is the kth output bit.

In the CVSD, $\Delta_{max} = 2$ dBm, which places the center of the coder dynamic range near -21 dBm, the central value of the three signal input levels used in the experiment.

2.3.2 ADM

The double integration version of CVSD was selected for evaluation after a large number of other delta modulators were simulated. The other delta modulators differed from CVSD in one or more of the following respects:

(*i*) An exponential expandor was used in the step-size feedback loop to produce the step size

$$\Delta'_{k+1} = \exp(\Delta_{k+1}),$$

where Δ_{k+1} is given by (2). This changes the adaptation from essentially additive to essentially multiplicative.

(*ii*) The most recent two bits rather than the most recent three bits were used to determine whether the step size would increase or decrease.

(*iii*) The signal feedback loop contained a double integrator rather than a single integrator.

A limited amount of speech material was processed to evaluate these modifications. Segmental s/n was measured for each delta modulator configuration and, altbough some modifications resulted in better performance than CVSD for certain input levels, none of them produced substantially better results either in terms of dynamic range or peak segmental s/n. However, to provide one delta modulation alternative

to CVSD, we chose the double integration ADM. Double integration is known to enhance performance significantly at higher sampling rates and to be essentially ineffective in 8-kHz DPCM (see Section 2.4). Our purpose here was to assess the effectiveness of a particular double integrator in a coder with 16-kHz sampling.

The double integrator in this coder is a second-order FIR filter that conforms to the block diagram in Fig. 4. The difference equations of the integrator are

$$u_{k+1} = y_k + H(1 - \alpha_1 - \alpha_2)b_k\Delta_k \tag{4}$$

$$y_{k+1} = \alpha_1 u_{k+1} + \alpha_2 u_k, \tag{5}$$

where

u_k is the decoder output,
y_k is the output of the encoder feedback loop,
$\alpha_1 = 1.38$, $\alpha_2 = -0.43$ are the filter coefficients, and
$H = 10$ is the gain.

The z-transform of the integrator is

$$\frac{\alpha_1 z^{-1}(1 - c_3 z^{-1})}{(1 - c_1 z^{-1})\,(1 - c_2 z^{-1})}, \tag{6}$$

where the integrator poles are related to the filter coefficients by

$$c_1 + c_2 = \alpha_1 \quad c_1 c_2 = -\alpha_2 \tag{7}$$

and the zero is

$$c_3 = -\alpha_2/\alpha_1. \tag{8}$$

The corresponding real poles and zero of the filter frequency response are

$$f_i = \frac{1}{2\pi T}\cos^{-1}\left[\frac{4c_i - c_i^2}{2c_i}\right], \tag{9}$$

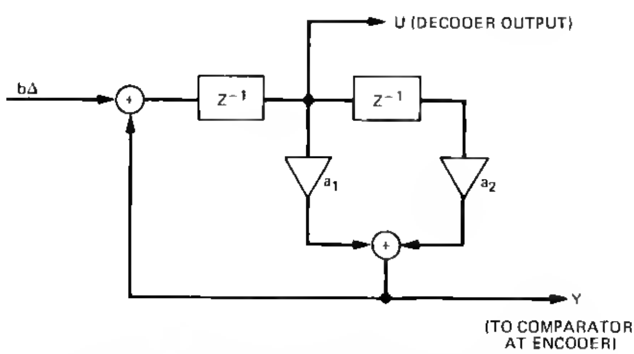

Fig. 4—Double integrator circuit for the ADM coder.

where $T = 1/16000$ s in a 16-kb/s delta modulator. With $\alpha_1 = \alpha_2 = 1.38$, -0.43, the pole frequencies are 200 and 2000 Hz, and the zero is at 3500 Hz, so that the integrator frequency response is approximately that shown in Fig. 5. In the ADM, $\Delta_{max} = -5$ dBm, which approximately centers the coder dynamic range at -21 dBm.

2.4 ADPCM

Figure 6 shows the block diagram of the ADPCM system. In an error-free environment, the primed quantities at the receiver are equal to the corresponding ones at the transmitter. In the encoder error sample $e(k)$ is generated as the difference between the input speech sample $s(k)$ and a predicted sample $\hat{s}(k)$. After quantization with 2 bits/sample, the prediction error at both receiver and transmitter is added to the predicted sample to give the reconstructed sample $r(k)$. The predicted sample $\hat{s}(k)$ is derived from the previous reconstructed one, $r(k-1)$, by a first-order transversal predictor:

$$\hat{s}(k) = 0.78r(k-1). \tag{10}$$

The coefficient 0.78 was computed by the usual mean-square error minimization technique[9] under the hypothesis of an overall s/n of about 10 dB.

The 2-bit coding of the prediction error is effected by means of the adaptive step size $\Delta(k)$, which is computed as proportional to a short-time estimate $\sigma(k)$ of the absolute magnitude of the quantizer input. The estimate $\sigma(k)$ is computed recursively from the decoded prediction error $d(k)$ so that no side information has to be transmitted. The

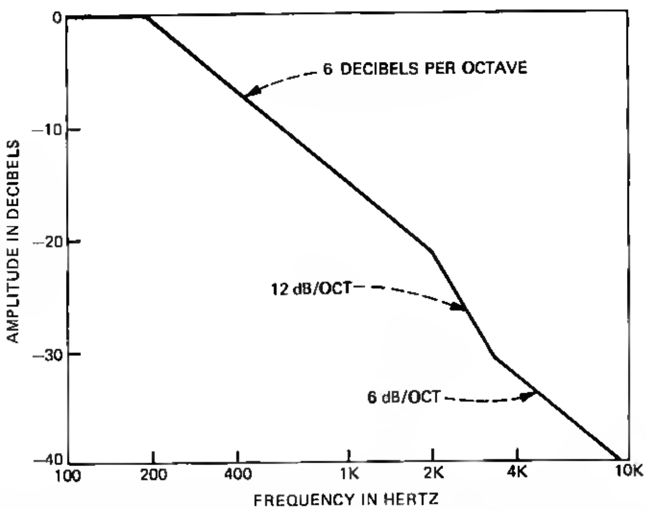

Fig. 5—Frequency response of the double integrator circuit.

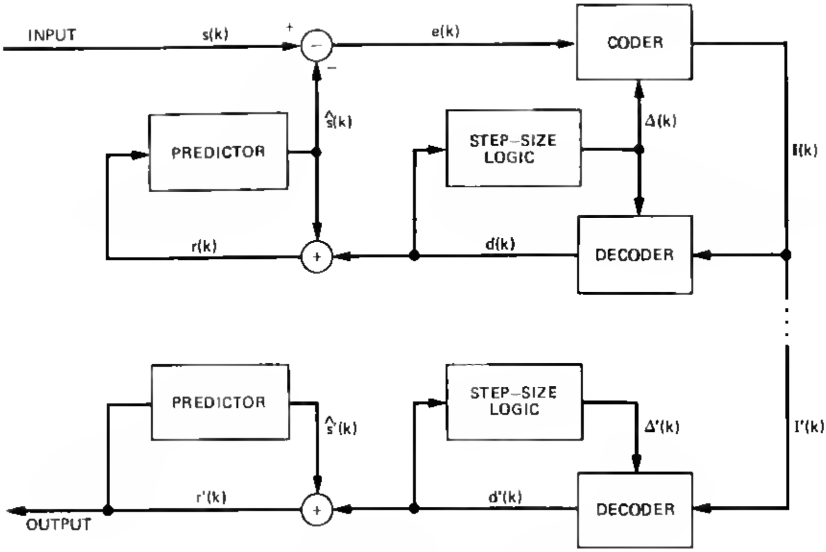

Fig. 6—Block diagram of the ADPCM coder.

entire adaptation process is summarized by the following two equations:

$$\Delta(k) = C\sigma(k) \tag{11}$$

$$\sigma(k) = \alpha\sigma(k-1) + (1-\alpha)\,|\,d(k-1)\,|\,. \tag{12}$$

In eq. (11), the parameter C, the load-constant, determines in the steady state the magnitude of the average step-size and hence the amount of granular noise and overload distortion. In eq. (12), the parameter α determines the speed of response of the adaptation algorithm to input level changes: a relatively small value of α produces fast response but an inaccurate estimate in steady state.

For simulating a practical implementation, the step size $\Delta(k)$ was constrained to assume values in the range $(\Delta_{min}, \Delta_{max})$ with

$$\frac{\Delta_{max}}{\Delta_{min}} = 256. \tag{13}$$

Values of C, α, and Δ_{min} were calculated by optimizing a prediction of the subjective opinion score, obtained from separate measures of granular noise and overload distortion. The integrator coefficient was $\alpha = 0.875$, corresponding to a time constant of 1 ms. The minimum step size which produced the same degradations at the high and low end of the input level range of interest was found to be −63 dBm. The maximum step size is −15 dBm, while the rms speech input level assumes values in the range −36 dBm to −6 dBm.

2.5 The sub-band coder

Sub-band coding is a waveform coding technique in which the speech band is partitioned into typically 4 or 5 sub-bands by bandpass filters. Each sub-band is then lowpass-translated to dc, sampled at its Nyquist rate, and then digitally encoded using adaptive PCM (APCM) encoding. By this process of dividing the speech band into sub-bands, each sub-band can be preferentially encoded according to perceptual criteria for that band. On reconstruction, sub-band signals are decoded and bandpass-translated back to their original bands. They are then summed to give a replica of the original speech signal.

A particularly attractive implementation of the sub-band coder in terms of hardware is based on an integer band sampling approach.[10-12] With this approach, the modulations to lowpass at the transmitter and to bandpass at the receiver are inherent in the sampling process. The implementation is illustrated in Fig. 7. Bandpass filters BP_1 to BP_N in the transmitter and receiver serve to partition the input speech into N sub-bands. The coders and decoders encode the sub-band signals and the multiplexer combines these digital signals and synchronizing bits into a single bit stream for transmission over the digital channel.

Table I shows the choice of bands and bit allocations used in the 16

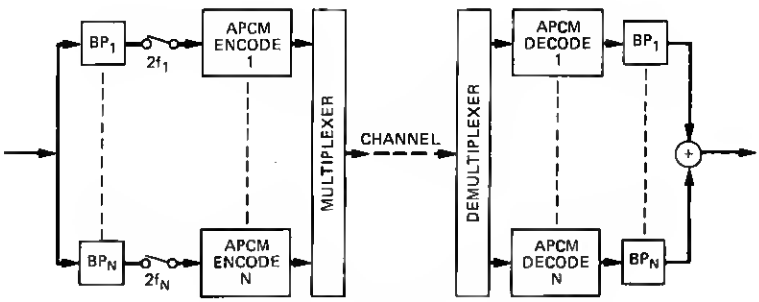

Fig. 7—Block diagram of the sub-band coder.

Table I—16 kb/s 5-band sub-band coder

Band	Band Edges (Hz)	Sampling Freq (Hz)	Min. Step-size (dB)	Bit Allocation	Kb/s
1	178–356	356	(Ref)	4	1.42
2	296–593	593	0	4	2.37
3	533–1067	1067	0	3	3.20
4	1067–2133	2133	−3	2	4.27
5	2133–3200	2133	−8	2	4.27
Sync					0.47
					16.00

kb/s coder. The coder is a 5-band design which was proposed in Ref. 11. Column 2 shows the frequency range covered by each sub-band. The bit allocation refers to the number of bits/sample used by the coders in each sub-band. As seen from the table, more accuracy is allowed for encoding the lower bands for reasons explained in Ref. 11.

The frequency range of the coder extends from 200 to 3200 Hz. A plot of the frequency response, shown in Fig. 8, reveals small notches at 1067 and 2133 Hz. These notches are due to the transition bands of the filters in bands 4 and 5. Subjectively, they are not very perceptible. Bands 1 to 3 are overlapped to avoid such notches at lower frequencies. The filters are sharp cutoff, 200 tap, FIR filters.

Column 4 in Table I contains the minimum step sizes of the APCM coders, expressed in decibels, relative to the minimum step size of band 1. This choice of minimum step sizes is different than that suggested in Ref. 11 and was found to give a better matching of the dynamic ranges of the sub-bands.

III. OBJECTIVE MEASUREMENTS

Four different objective measurements were made on the waveform coders. They are conventional signal-to-noise ratio, *SNR*, two types of segmental signal-to-noise ratios, *SEG*1 and *SEG*2, and an LPC spectral distance measure, D. In addition, D was used to evaluate the performance of the LPC vocoder and the tandem connections of the waveform coders and the LPC vocoder. In this section we briefly define each of these objective measures.

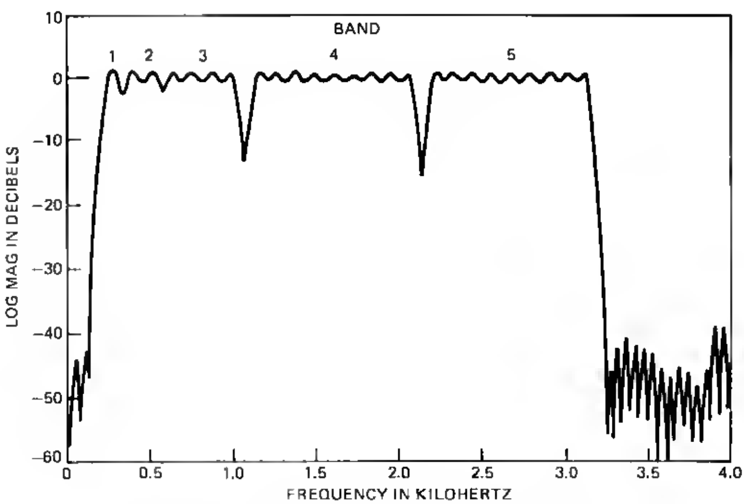

Fig. 8—Frequency response of the sub-band coder.

3.1 Conventional SNR

The most commonly used measure of performance of digital coders has been the conventional signal-to-noise ratio evaluated over an utterance of speech. The speech power is defined as

$$\hat{p} = \sum_m x^2(m), \tag{14}$$

and the noise power is defined as

$$\hat{n} = \sum_m (x(m) - y(m))^2, \tag{15}$$

where $x(m)$ and $y(m)$ are the input and output signals of the coder, respectively, and the summations in (14) and (15) are taken over the entire speech utterance. The conventional s/n is then defined as

$$\text{SNR} = 10 \log(\hat{p}/\hat{n}). \tag{16}$$

In measuring the input and output signals of the coders, it is generally desirable to compensate for the effects of lowpass or bandpass filtering. This is done by the arrangement shown in Fig. 9. The input speech signal $s(m)$ is coded to form the output speech signal $y(m)$. It is also filtered with the same filters used in the coder to generate a compensated reference signal $x(m)$ which is used as the input signal in (14) and (15). SNR is, therefore, strictly a measure of coder distortions and is not affected by bandlimiting or group delay in the filters.

3.2 SEG1

While SNR is the most widely used criterion in measuring coder distortion, it has also long been known that in many situations it does not correlate well with subjective performance.[13] Another definition of signal-to-noise ratio, however, recently proposed by Noll,[3] does appear to correlate better with subjective performance. This measure is based on s/n measurements made over segments of speech which are typically about 20 ms in duration. An average over all of the segments in the speech utterance is then taken to obtain a composite measure of

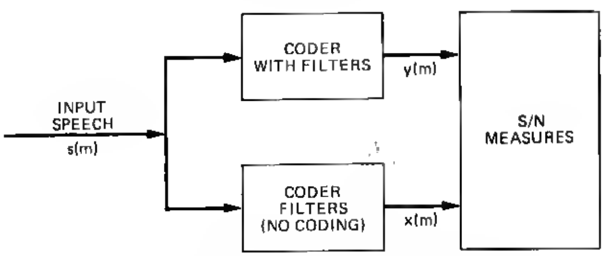

Fig. 9—Circuit for measuring signal-to-noise ratios.

performance for the entire utterance. If $(s/n)_i$ corresponds to the signal-to-noise ratio in decibels for a segment, i [computed in the same manner as in (16)], the segmental s/n (SEG1) is then defined as

$$\text{SEG1} = \frac{1}{N}\sum_{i=1}^{N} (s/n)_i \quad \text{(dB)}, \tag{17}$$

where it is assumed that there are N 20 ms segments in the speech utterance.

Problems arise in this definition of segmental s/n when intervals of silence exist in the speech utterance. In segments where the input signal $x(n)$ is essentially zero, any slight noise will give rise to large negative $(s/n)_i$, and these segments may unduly dominate the average in (17). To prevent this anomaly, we first identify those segments which correspond to silence and exclude them from the average in (17). This is achieved by means of a simple threshold. Let $\hat{p}_i$ represent the (average) speech energy in a segment, i, so that

$$\hat{p}_i = \frac{1}{K}\sum_{m=1}^{K} x^2(m), \tag{18}$$

where K corresponds to the number of speech samples in the segment. Then the segment will be included in the computation of SEG1 in (17) if its energy exceeds a threshold T, i.e., if $\hat{p}_i > T$. If it does not exceed this threshold, it is not included in the average in (17). Furthermore, to prevent any one segment from dominating the average we also limit the value of $(s/n)_i$ to a range of −10 to +80 dB. That is, $-10 \leqslant (s/n)_i \leqslant 80$ dB. In computer simulations, the 16-bit wordlength admitted signal levels in the range ±32767 and we set T to 900, corresponding to −55 dBm.

3.3 SEG2

The definition of this measure is

$$\text{SEG2} = \frac{1}{N}\sum_{i=1}^{N} 10 \log_{10}(1 + \hat{p}_i/\hat{n}_i) \quad \text{(dB)}, \tag{19}$$

where $\hat{p}_i$ is the signal power in segment i and $\hat{n}_i$ is the noise power in segment i. They are defined (on segments) according to eqs. (14) and (15), respectively.

Unlike SEG1, SEG2 does not have any thresholds. It is self-limiting to a lower value of 0 dB due to the addition of the constant 1 inside the logarithm. As in the SEG1 measure, SEG2 uses 20 ms segments.

3.4 LPC distance measure

The fourth objective measure was the LPC distance proposed by Itakura.[14] The distance between two frames of speech with LPC coef-

ficient vectors $\mathbf{a}$ and $\hat{\mathbf{a}}$, and with autocorrelation matrices $\mathbf{V}$ and $\hat{\mathbf{V}}$ is defined as

$$D_1 = d(\mathbf{a}, \hat{\mathbf{a}}) = \log\left[\frac{\mathbf{a}\hat{\mathbf{V}}\mathbf{a}^t}{\hat{\mathbf{a}}\hat{\mathbf{V}}\hat{\mathbf{a}}^t}\right], \tag{20}$$

where $\mathbf{a}$ and $\hat{\mathbf{a}}$, are $(p + 1)$ component vectors and $\mathbf{V}$ and $\hat{\mathbf{V}}$ are $(p + 1) \times (p + 1)$ matrices, where p is the order of the LPC analysis.

D_1 is a measure of the distance between frames of speech. This distance, however, does not satisfy exactly all the properties of a true distance metric, i.e.,

$$d(\mathbf{a}, \hat{\mathbf{a}}) \neq d(\hat{\mathbf{a}}, \mathbf{a}). \tag{21}$$

However, for cases when $d(\mathbf{a}, \hat{\mathbf{a}})$ is small, the inequality of eq. (21) is almost an equality. To compensate for this lack of symmetry, it is convenient to define a second distance, D_2, as

$$D_2 = d(\hat{\mathbf{a}}, \mathbf{a}) = \log\left[\frac{\hat{\mathbf{a}}\mathbf{V}\hat{\mathbf{a}}^t}{\mathbf{a}\mathbf{V}\mathbf{a}^t}\right] \tag{22}$$

and an average distance between the two frames is now given by

$$D = \frac{D_1 + D_2}{2}. \tag{23}$$

Equation (23) defines a true distance metric which can be used to measure the distance (dissimilarity) between two frames of speech. It can readily be shown[15] that either D_1 or D_2 can be expressed in terms of spectral differences between the LPC models for the two frames of speech.

D_1 and D_2 were measured for every utterance used in the tests to be described later. They were measured on a frame-by-frame basis and averaged across the entire utterance to give an overall LPC distance between the original and processed version of a sentence. Figure 10 shows the system used to measure LPC distance for a single coder. The box labeled delay was used to compensate any delay inherent in the coder, and the bandpass filters were used both to eliminate out-of-

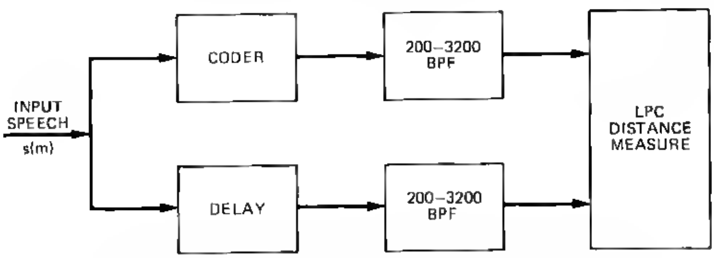

Fig. 10—Circuit for measuring LPC distance measures.

band quantization noise generated in the coder and to guarantee that the bandwidths of both the original and coded utterances were the same.

3.5 Comparison of SEG and D

Figures 11 and 12 show a series of plots for two of the utterances used in the experiments (encoded with the ADPCM coder). Part **a** of each figure shows the rms energy of the utterance as a function of time (frame number), part **b** of each figure shows the segmental s/n versus frame number, and part **c** of each figure shows the LPC distances

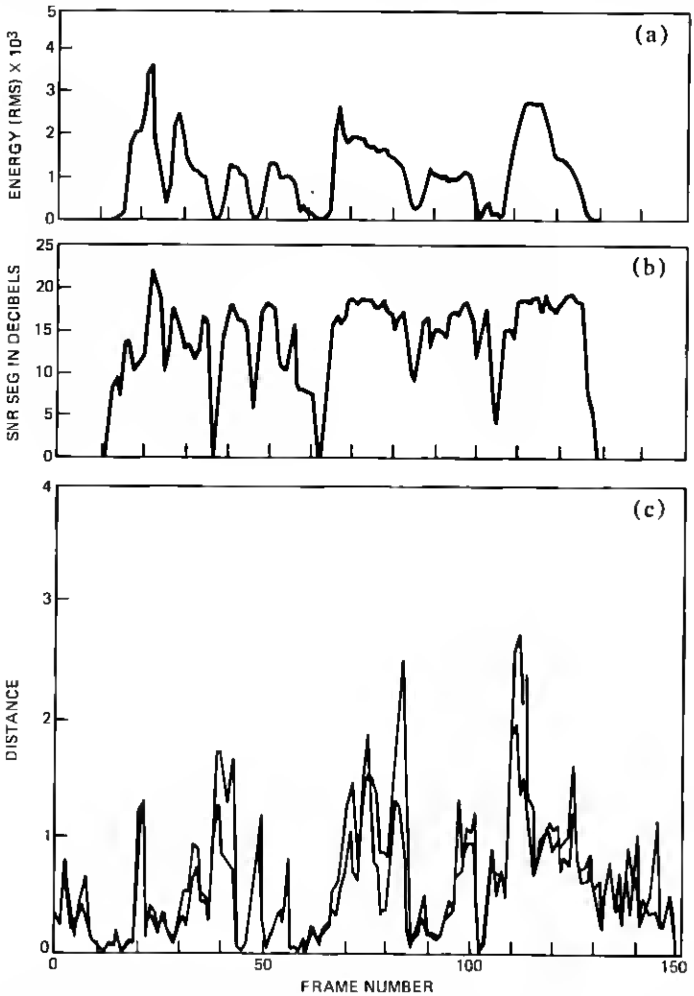

Fig. 11—Objective measurements as a function of time for utterance *A*. (a) rms energy of the input signal. (b) Segmental s/n-*SEG*1. (c) LPC distance.

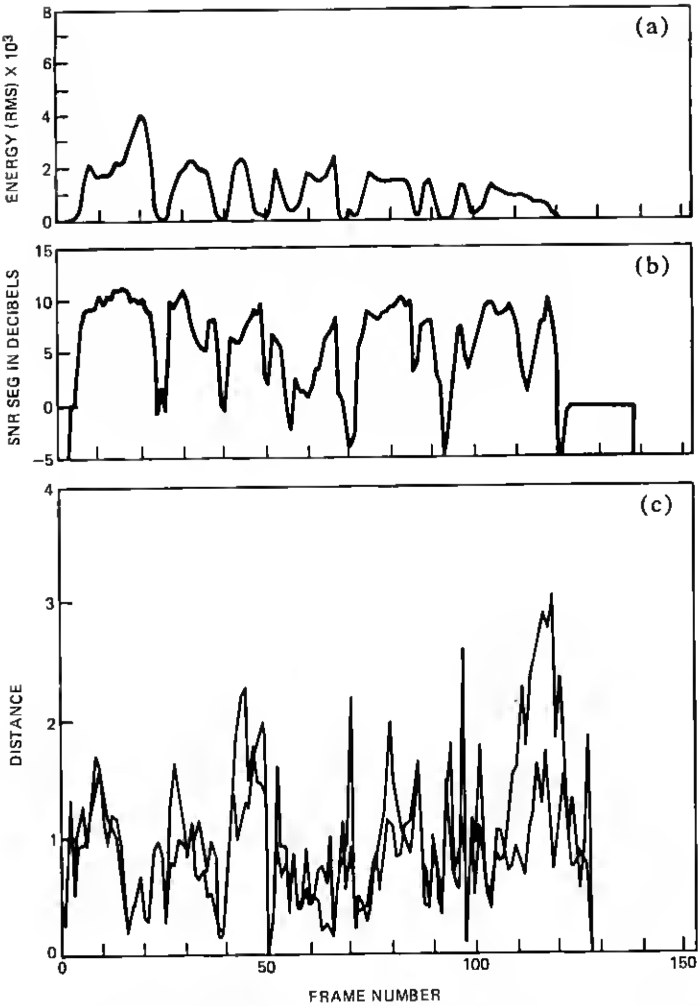

Fig. 12—Objective measurements as a function of time for utterance *B*. (a) rms energy of the input signal. (b) Segmental s/n-*SEG*1. (c) LPC distance.

(botb D_1 and D_2) versus time. The utterance of Fig. 11 had an average LPC distance of about 0.60, whereas the utterance of Fig. 12 had an average LPC distance of 0.97. It can be seen in both figures that most of the time $D_1 \approx D_2$; however when either D_1 or D_2 was large, the differences between D_1 and D_2 were often significant. It can also be seen in these figures that the LPC distance and the segmental s/n do not measure similar types of distortion—i.e., when the segmental s/n is small (indicating temporal distortion of tbe waveform) the LPC distance is not necessarily large (indicating spectral distortion of the signal). Finally, it can be seen that a large degree of variation (on a

frame-by-frame basis) occurs with both the segmental s/n and the LPC distance. Thus, a single number which characterizes the "distortion" across the entire utterance may have little meaning in many cases.

IV. EXPERIMENTAL DESIGN

4.1 Circuit conditions

The experiment tested 37 different communication circuits, each characterized by three parameters: direction, coder, and level. There were three directions: (*i*) single link with a waveform coder or vocoder alone, (*ii*), LPC-to-waveform, as in Fig. 1, (*iii*) waveform-to-LPC as in Fig. 2. There were five different single links, four with waveform coders and one with a vocoder. Each waveform coder was tested with speech at three different input levels, −36 dBm, −21 dBm, and −6 dBm. The corresponding settings of the gain parameter, G, were 0.178, 1.00 and 5.62, respectively. The speech level at the vocoder input was always −21 dBm. Thus, there were 13 single link configurations, in all. Each of the other two directions had 12 circuit configurations, comprising all combinations of four waveform coders and three input levels.

4.2 Speech material

For this experiment, a substantial digital speech library was prepared. Four talkers, two male and two female, read 40 different sentences, 2 to 3 seconds long, each talker reading from a different phonetically balanced list. The talkers were seated in a sound-proof booth and spoke into a high-quality dynamic microphone. The amplified microphone signal was lowpass-filtered at 3.2 kHz, sampled and converted into digital form by a 16-bit A/D converter operating at 8-kHz sampling frequency and finally written onto a magnetic disk. All the sentences were digitally equalized to the mean power level of −21 dBm.

For each of the 37 circuit conditions, sentences spoken by each of the four talkers were processed, generating a total of 148 stimuli. Different sentences were used in each case so that in the set of 148 stimuli the same sentence was never heard twice. With this format, we speculate that intelligibility of the processed speech played an important role in determining quality judgments. In tests containing a few sentences, presented many times, each sentence becomes recognizable to subjects even in conditions severe enough to make it quite unintelligible at first hearing. It is our hypothesis that, in such tests, there is a lower correlation between intelligibility and subjective quality than in the tests reported here.

4.3 Procedure

The 148 stimuli were recorded in different random orders on 4 analog tapes. Twenty-two students from the junior and senior classes

of local high schools served as paid subjects. They listened to the processed speech monaurally over Pioneer SE 700 earphones at 80 dB SPL while seated in a double-walled sound booth with frequency-weighted room noise introduced at a level of 50 dBA. The total listening time for each group of subjects was about 30 minutes, with a short break after the 80th sentence. After each stimulus, the subjects had 4 seconds for recording their judgments. They were instructed to rate the quality of the stimulus by checking on their answer sheets a value between 1 and 9, using 1 for the worst conditions, 9 for the best ones, and intermediate numbers for intermediate qualities. Before the actual test, the subjects listened to 12 practice sentences, different from those used in the experiment, spanning the range of quality in the experiment.

V. SPEECH QUALITY RESULTS

Variability in the subjective and objective measures of quality of the 148 processed speech samples can be attributed to several (variable) sources, namely:

(*i*) The "direction" of the circuit: LPC-to-waveform coder, waveform-coder-to-LPC, or single link.
(*ii*) The waveform coder.
(*iii*) The speech level at the input to the coder.
(*iv*) The talker.
(*v*) The sentence.
(*vi*) The listener (subjective data only).
(*vii*) Inconsistency of each listener (subjective data only).

We are primarily interested in how the first two variables, circuit direction and waveform coder, influence quality. Inferences about these variables would be simple if they accounted for most of the variance in the data or if they did not interact substantially with the other variables. Unfortunately, neither of these conditions is met by our data, and many of our inferences about circuit direction and waveform coder will be more qualified than we would like them to be.

5.1 Subjective data

5.1.1 Listeners

The amount of listener agreement was fairly low relative to other speech quality experiments.[3,13] For each pair of listeners, we computed the correlation coefficient of the 148 ratings. The median of the correlations was only 0.49. The 25th and 75th percentiles were 0.41 and 0.60, respectively. With respect to the 148 mean ratings (averaged over the 22 listeners), individual listener correlations ranged from 0.50 to 0.85, which suggests that no subject was very idiosyncratic in his ranking of stimulus conditions.

Figure 13 gives plots of the rating scores of each of the 22 subjects for the LPC system alone and for each of the four talkers. The large variability among subjects is readily seen. For example, for talker 3 the average rating was 7.3. However, two subjects gave this circuit rating of 1 (the lowest possible), whereas 13 subjects gave it a rating of 8 or 9 (the highest possible). Similar variability was found in the scores for almost every test condition.

The 148 listener averages are presented in Table II, where we also provide aggregates of these averages across input level and talker. The aggregated mean values show the overall effects of circuit direction and waveform coder.

5.1.2 Sentences

In many subjective testing experiments, listeners hear one or a few sentences repeatedly. To achieve closer conformity to practical communication situations, a different sentence for each stimulus condition was used. A disadvantage of this design is the lack of any control for or means of testing the effect of sentence content on the quality measures. The variability due to sentences appears in and enhances the experimental error; i.e., the variance that cannot be accounted for in statistical analyses.

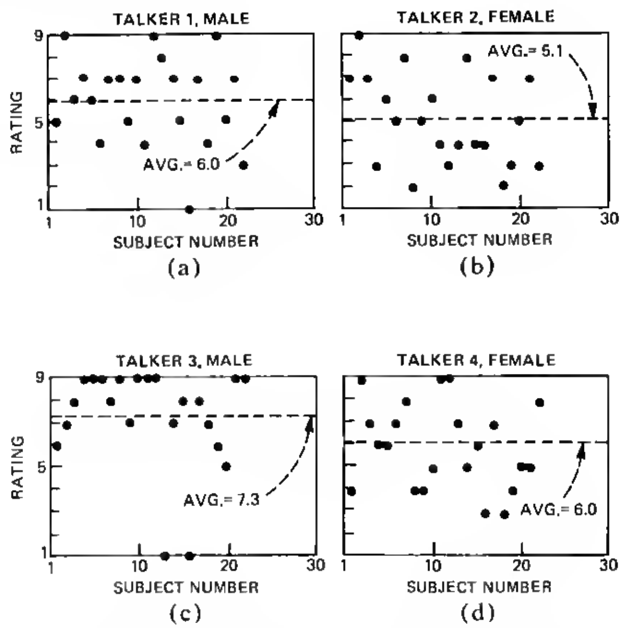

Fig. 13—Rating scores as a function of subject for the individual LPC circuit for each of the four talkers.

5.1.3 Talker effects

Averaged over all 37 circuit conditions (combinations of input level, coder, and direction), the ratings of speech from the two male talkers were 4.98 and 4.94. The averages for the two females were 3.99 and 4.00. A three-way analysis of variance (listener by talker by circuit condition) revealed a very significant talker effect. Clearly, this effect is predominantly due to listeners giving lower ratings to distorted female speech than to distorted male speech. However, there is also a substantial talker-circuit interaction, indicating that differences in ratings of male and female speech are by no means uniform across experimental conditions. (In fact, with fairly low distortion as in sub-band coding in a single link, the male and female averages are virtually the same—7.13 and 7.20, respectively.) This nonuniformity is evident in Table II which also reveals that, although the overall ratings of the two males are virtually identical, there are substantial differences from condition to condition in the ratings of the male voices, and likewise for the female voices.

5.1.4 Input level

The step sizes of the waveform coders were adaptable over a range of 44 dB (for the CVSD) or 48 dB (for the other three coders). With the rms input level varying over a range of 30 dB and individual sounds within a sentence exhibiting a wide range of levels, the weak sounds of the low-level signals were subject to greater-than-average granular quantizing noise, while the strong sounds of the high level sentences were susceptible to overload. The maximum and minimum step sizes of each waveform coder were chosen with the aim of centering the dynamic range of subjective quality in the −15 to +15 dB range of input levels.

Table III shows that this design effort was entirely successful with the CVSD and ADPCM coders in which the dynamic range of subjective performance is exactly symmetric around the 0-dB input level. In the SBC and ADM coders, the overload distortion of the +15 dB input level was less harmful subjectively than the granular noise produced with the input set at −15 dB. In these coders, a better balance of granularity and overload would have been achieved with lower minimum step sizes.

5.1.5 Coder and direction

We have used the Tukey HSD criterion[16] to evaluate the relative merits of the 13 communication system configurations listed in Fig. 14. Figure 14a shows, for each circuit direction, groupings of coders for which the null hypothesis cannot be rejected at the 0.05 level. In all cases, SBC is superior to any of the other waveform coders. In the LPC

Table II—Average subjective ratings (over 22 listeners)

		Single Link					LPC → Waveform				Waveform → LPC			
		SBC	CVSD	ADPCM	ADM	LPC	SBC	CVSD	ADPCM	ADM	SBC	CVSD	ADPCM	ADM
M1	−15	6.8	6.3	5.4	4.9		4.5	3.2	2.9	3.1	4.7	2.5	5.7	3.9
	0	6.3	6.7	7.2	7.0	6.0	4.3	4.9	4.3	4.5	4.5	2.9	5.1	4.0
	+15	7.5	6.2	5.9	6.8		5.0	4.8	4.0	4.5	4.5	3.8	5.5	4.2
	Ave (level)	6.9	6.4	6.2	6.2		4.6	4.3	3.7	4.0	4.6	3.0	5.4	4.0
M2	−15	7.6	7.0	6.5	5.5		5.1	3.9	3.8	3.7	4.8	4.3	3.5	3.6
	0	7.6	6.9	5.6	5.9	7.3	5.7	2.9	4.0	3.5	6.1	3.6	4.7	3.5
	+15	6.9	5.5	5.6	5.9		6.1	3.1	2.5	4.5	5.0	3.8	4.0	3.4
	Ave (level)	7.4	6.5	5.9	5.8		5.6	3.3	3.5	3.9	5.3	3.9	4.1	3.5
Ave (M1, M2, level)		7.1	6.5	6.0	6.0	6.6	5.1	3.8	3.6	4.02	4.9	3.5	4.8	3.8
F1	−15	7.0	4.8	3.5	3.2		3.8	2.3	2.1	3.1	2.0	3.9	2.4	2.0
	0	7.9	6.2	5.4	4.5	5.1	5.3	2.8	3.7	2.2	1.6	3.3	2.3	2.7
	+15	7.8	3.9	4.8	6.5		6.5	2.9	2.5	3.1	6.9	3.0	3.4	3.3
	Ave (level)	7.6	5.0	4.5	4.7		5.2	2.7	2.8	2.8	3.5	3.4	2.7	2.7
F2	−15	5.7	5.8	4.6	4.1		3.4	3.6	3.5	2.9	4.2	1.8	3.3	2.8
	0	7.4	4.7	3.4	6.0	6.0	4.8	2.6	3.2	4.9	4.3	1.9	3.2	2.5
	+15	7.4	5.8	4.5	4.6		3.9	3.0	2.1	3.1	4.5	3.2	2.5	2.7
	Ave (level)	6.8	5.4	4.2	4.9		4.0	3.1	3.0	3.6	4.3	2.3	3.0	2.7
Ave (F1, F2, level)		7.2	5.2	4.3	4.8	5.6	4.6	2.9	2.9	3.2	3.9	2.8	2.9	2.7
Ave (Talker, level)		7.2	5.8	5.2	5.4	6.1	4.9	3.3	3.2	3.6	4.4	3.2	3.8	3.2

Table III—Average subjective ratings (over listeners, talkers and direction)

		Coder			
		SBC	CVSD	ADPCM	ADM
Level	−15 dB	5.0	4.1	3.9	3.6
	0 dB	5.5	4.1	4.3	4.3
	+15 dB	6.0	4.1	3.9	4.4

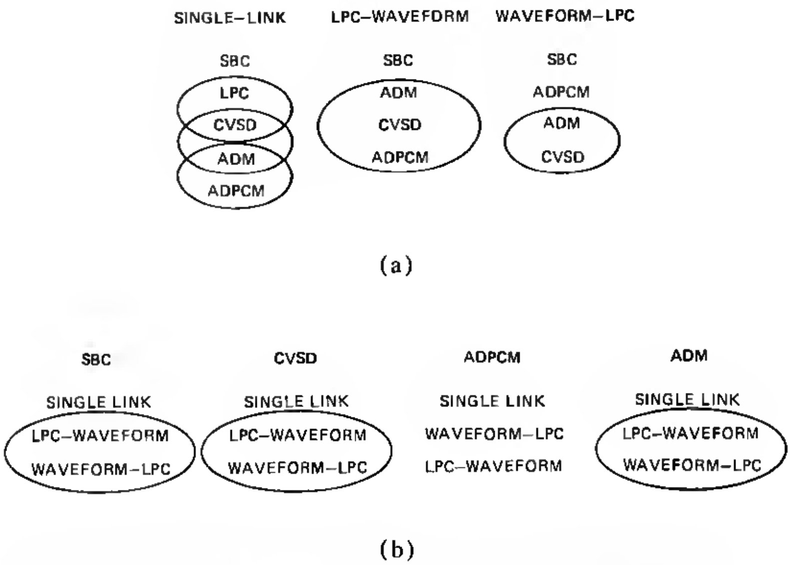

Fig. 14—Relative subjective quality of coding systems. Circles indicate that it is impossible to reject the hypothesis that the coders have the same quality.

→ waveform circuits, ADM, CVSD, and ADPCM have essentially the same performance. In the waveform → LPC direction, ADPCM is better than ADM and CVSD, which exhibit essentially the same quality.

Figure 14b shows the equivalent groupings across direction. The salient inferences from these groupings is that, for each waveform coder, the single link substantially outperforms either of the tandem connections. The two tandem directions have essentially the same quality when SBC, CVSD, or ADM is the waveform coder. The ADPCM → LPC tandem is significantly better than the LPC → ADPCM tandem.

5.2 Objective measurements of quality

Results of the objective measurements discussed in Section III are presented in Tables IV and V. Table IV gives results for the performance of the single-link circuits in terms of *SNR*, *SEG*1, *SEG*2, and LPC

Table IV—Objective measurements of single link coders

Level	Coder	SNR	SEG1	SEG2	D	Level	Coder	SNR	SEG1	SEG2	D
	Talker: M1						Talker: F1				
−15	SBC	13.2	11.9	8.7	0.82	−15	SBC	17.7	15.3	12.3	0.53
0	SBC	14.4	13.4	8.8	0.50	0	SBC	14.4	13.3	10.5	0.51
+15	SBC	13.8	10.2	8.6	0.80	+15	SBC	13.2	14.9	12.0	0.34
−15	CVSD	12.1	10.2	9.1	0.65	−15	CVSD	14.8	12.1	11.6	0.84
0	CVSD	8.4	12.1	11.0	0.54	0	CVSD	17.5	15.2	13.8	0.65
+15	CVSD	3.5	8.3	9.8	0.54	+15	CVSD	8.2	11.8	12.3	0.55
−15	ADPCM	12.6	12.8	10.3	0.54	−15	ADPCM	16.2	16.2	13.9	0.61
0	ADPCM	10.1	12.3	12.0	0.34	0	ADPCM	16.5	14.0	13.9	0.50
+15	ADPCM	3.6	10.6	12.0	0.37	+15	ADPCM	8.2	12.6	13.1	0.58
−15	ADM	14.6	10.5	9.5	0.64	−15	ADM	16.5	14.7	12.6	0.89
0	ADM	12.3	13.2	11.1	0.47	0	ADM	18.4	16.5	14.9	0.70
+15	ADM	4.7	10.7	12.2	0.47	+15	ADM	10.4	12.7	13.0	0.48
—	LPC				0.31	—	LPC				0.38
	Talker: M2						Talker: F2				
Level	Coder	SNR	SEG1	SEG2	D	Level	Coder	SNR	SEG1	SEG2	D
−15	SBC	12.3	10.6	7.6	0.76	−15	SBC	14.6	14.1	12.0	0.60
0	SBC	13.8	14.5	10.8	0.38	0	SBC	13.4	14.1	11.5	0.33
+15	SBC	12.3	14.1	11.1	0.27	+15	SBC	12.2	13.8	12.1	0.28
−15	CVSD	12.0	9.9	8.1	0.81	−15	CVSD	12.8	11.5	10.0	0.74
0	CVSD	9.4	9.6	9.4	0.43	0	CVSD	11.1	14.9	12.5	0.62
+15	CVSD	4.4	9.9	11.5	0.43	+15	CVSD	4.3	11.3	12.1	0.40
−15	ADPCM	12.8	13.0	10.5	0.50	−15	ADPCM	14.5	14.8	12.5	0.64
0	ADPCM	12.0	13.8	13.4	0.45	0	ADPCM	14.3	14.9	14.5	0.38
+15	ADPCM	5.1	11.3	13.0	0.47	+15	ADPCM	4.5	11.2	12.6	0.67
−15	ADM	14.1	10.5	8.2	0.73	−15	ADM	16.5	12.5	10.5	0.66
0	ADM	11.6	12.1	11.5	0.38	0	ADM	17.8	16.4	14.7	0.58
+15	ADM	8.8	10.7	11.6	0.35	+15	ADM	8.9	13.1	12.6	0.51
	LPC				0.35		LPC				0.47

Table V—Overall LPC distances for tandem links

First Link	Second Link	Talker: M1	Talker: F1	Talker: M2	Talker: F2
LPC	SBC	1.18	0.93	0.83	0.81
LPC	SBC	0.20	0.73	0.672	0.66
LPC	SBC	0.61	0.57	0.46	0.56
LPC	CVSD	0.91	0.92	0.88	1.10
LPC	CVSD	0.66	1.00	0.54	0.91
LPC	CVSD	0.60	0.66	0.64	0.83
LPC	ADPCM	0.79	0.90	0.69	1.09
LPC	ADPCM	0.71	0.75	0.53	0.68
LPC	ADPCM	0.61	0.72	0.58	0.76
LPC	ADM	0.90	1.40	0.82	0.91
LPC	ADM	0.68	0.80	0.68	0.77
LPC	ADM	0.59	0.64	0.60	0.90
SBC	LPC	1.50	1.08	0.86	0.93
SBC	LPC	1.05	0.88	0.78	0.78
SBC	LPC	0.71	0.57	0.61	0.62
CVSD	LPC	0.91	0.86	0.98	1.21
CVSD	LPC	0.62	0.75	0.63	0.84
CVSD	LPC	0.62	0.82	0.56	0.89
ADPCM	LPC	0.75	0.76	0.71	0.79
ADPCM	LPC	0.63	0.80	0.67	0.87
ADPCM	LPC	0.51	0.74	0.60	0.76
ADM	LPC	0.79	0.71	0.86	1.06
ADM	LPC	0.58	0.81	0.64	0.80
ADM	LPC	0.47	0.62	0.56	0.69

distance D for each of the four talkers used in the experiment. Due to the large variability of the objective measures across talkers and sentences (a different sentence was used for each condition), it is difficult to make meaningful comparisons across conditions. A similar variability was observed for the objective measurements across individual coders in the tandem links.

Table V gives results for LPC distance for the overall tandem links. Again, a large variability is seen across conditions due to the different sentences used for each measurement.

5.3 Relationship of subjective and objective measures

5.3.1 Correlations

Previous studies have demonstrated the inadequacy of *SNR* as an indicator of subjective quality and have pointed to segmental signal-to-noise ratio and to LPC distance metrics as more promising measures. In the present experiment, the diversity of speech material and of signal-processing approaches exceed those of previous studies, and thus the merits of single measures and combinations of measures as subjective quality indicators are tested more critically than ever before.

Table VI shows correlations of average rating with each of the objective measures. The subscripts A, B, and AB, appended to *SNR*, *SEG*1, *SEG*2, and D, refer to measures taken on the first link of a tandem circuit (or the entire single-link circuit), the second link of a tandem circuit, and the overall circuit, respectively.

Table VI indicates that the diversity of conditions either eliminates or dilutes the value of each of the measures as a predictor of speech quality. The table gives correlations of average rating (over 22 subjects) with each one of the objective measures. There are nine objective measures; 3 s/n's and one LPC distance for each half of a tandem connection, and the overall LPC distance. Except for D_A, the LPC distance of the first link, and D_{AB}, the overall LPC distance, none of the measures is applicable to all conditions. (For example, s/n is measured only in the first link in the single-link and waveform-to-LPC circuits. It is measured only in the second link in the LPC-to-waveform circuits.) In addition to the correlation, the table shows the number of data points used in the computation and the significance (two-tailed) of the null hypothesis that the coefficient is zero.

It should be noted that, for all talkers, the only statistic for which the null hypothesis can be rejected at the 0.01 level is D_{AB}, the overall distance. The two-tailed significance level for $SEG2_A$ is 0.001, but the correlation is negative. Surely a one-tailed test applies here, and the null hypothesis cannot be rejected. Computing correlations for ratings of male and female talkers separately, we see the same situation, except that $SEG1_B$ is significant at the 0.01 level as a predictor of male speech quality on the LPC-to-waveform tandems.

Table VI—Correlations of average ratings with objective measures

	All Talkers			Male Talkers			Female Talkers		
	Corr.	No. of Conditions	Signif.	Corr.	No. of Conditions	Signif.	Corr.	No. of Conditions	Signif.
SNR_A	−0.115	96	0.1	−0.038	48	0.4	0.014	48	0.5
$SEG1_A$	−0.175	96	0.04	−0.043	48	0.4	0.046	48	0.4
$SEG2_A$	−0.309	96	0.001	−0.196	48	0.09	−0.167	48	0.1
D_A	−0.089	148	0.1	0.233	74	0.02	−0.193	74	0.05
SNR_B	0.087	48	0.3	0.254	24	0.1	0.382	24	0.03
$SEG1_B$	0.072	48	0.3	0.510	24	0.005	0.381	24	0.03
$SEG2_B$	−0.123	48	0.2	0.318	24	0.06	0.031	24	0.4
D_B	−0.149	96	0.07	−0.056	48	0.4	−0.086	48	0.3
D_{AB}	−0.590	148	0.001	−0.415	74	0.001	−0.709	74	0.001

The poor correlations of practically all s/n measures with subjective quality has led us to abandon all of them as performance indicators of the tandem circuits and to focus our attention on the LPC distance measures.

5.3.2 Prediction of subjective quality

Working with the LPC distance measure for the first link of a tandem connection, D_A, the distance measure for the second link, D_B, and D_{AB}, the overall distance measure, linear regression procedures, were applied to find formulas for predicting the average ratings, $\bar{R}$, of the 148 circuit conditions. The best linear combination of the three distances was

$$\bar{R} = -5.48D_A - 6.47D_B + 2.52D_{AB} + 7.38. \tag{24}$$

The standard deviation of the 148 mean ratings was 1.55 units on the 9-point scale and the standard error of this regression was 1.10. The proportion of variance accounted for is thus 51 percent, and the multiple correlation coefficient is 0.712.

The prediction accuracy can be improved somewhat by accounting for the fact that ratings and LPC distances are related differently for male and female talkers. We have done so by introducing a new variable, M, where $M = 1$ for male talkers and $M = 0$ for female talkers. Introducing M to the regression, we have

$$\bar{R} = -4.99D_A - 5.98D_B + 2.14D_{AB} + 0.48M + 6.85. \tag{25}$$

Here the standard error is 1.08, i.e., 53 percent of the variance is accounted for and the multiple correlation coefficient is 0.727.

Various transformations of the distance data were also studied and a simple log transform proved useful in regression equations. We define the transform variables

$$L_A = \ln(D_A); \qquad L_{AB} = \ln(D_{AB})$$

and

$$\begin{aligned} L_B &= \ln(D_B) \text{ in tandem circuits and} \\ &= -4.0 \text{ in single-link circuits.} \end{aligned}$$

The value -4.0 has been chosen empirically. (It corresponds to a distance of 0.018. The lowest measured distance was 0.21, which was observed for several sentences processed by LPC.)

Using the log-transformed distances, the regression equations corresponding to (24) and (25) are

$$\bar{R} = -1.55L_A - 0.785L_B - 0.211L_{AB} + 1.59 \tag{26}$$

and

$$\bar{R} = -1.34L_A - 0.782L_B - 0.0622L_{AB} + 0.643M + 1.51. \quad (27)$$

The standard error of (26) is 1.02, which accounts for 58 percent of the variance in average ratings and the multiple correlation coefficient is 0.760. The corresponding statistics for (27) are 0.973, 62 percent, and 0.785.

VI. DISCUSSION

These data analyses allow us to make generalized statements in answer to the three questions posed in Section 1.2. Owing to the interactions in the data, there are specific exceptions to many of the general conclusions of the following subsections.

6.1 Quality of tandem connections

A strong conclusion of the study is that any tandem connection of the vocoder is substantially worse than either of the two corresponding single links. Although we did not attach descriptive adjectives to rating categories, we have the impression that ratings below about 4.0 reflected degradations severe enough to render a circuit inadequate for effective communication.

In our judgment, the results of this experiment strongly suggest that a tandem connection involving any of the three differential waveform coders (CVSD, ADPCM, or ADM) is inadequate. It appears that the LPC-SBC tandem could provide reasonable communication in many circumstances, but that the SBC-LPC tandem is of marginal use.

6.2 Alternatives to CVSD

Only the sub-band coder, which is substantially more complicated, offers significantly better performance than CVSD over all circuit conditions, talkers, and input levels. ADM, a double integration version of CVSD, has the same subjective quality (within the bounds of experimental error) and ADPCM is better than CVSD in one tandem direction, equal to CVSD in the other tandem direction and worse than CVSD in the single link configuration. The ADPCM coder was designed by extrapolating, to 16 kb/s, results of an experiment involving 24 kb/s and 32 kb/s coders. The result of this design optimization was a coder that adapts somewhat more slowly than ADPCM coders used elsewhere. It may be that higher quality could be obtained with a faster adaptive quantizer in the ADPCM coder.

6.3 Objective measures

The wide variety of circuit conditions and speech material either destroyed or strongly diluted the value of the objective measures as indicators of speech quality. With the wide range of input levels, the outputs of differential waveform coders contained various types of

additive noise and signal distortion. Meanwhile, the sub-band coder and LPC each have their own peculiar distortions; a reverberant effect and a mechanical buzziness, respectively. The presence of all these impairments in the single link circuits and their combinations in the tandem circuits together present a diversity of quality that would be very hard to describe with a single measure.

While the wide range of circuit conditions produces great subjective variability, the variety of speech material seems to have a strong effect on the objective measures. We speculate that sentence-to-sentence fluctuation in objective measures is greater than that of corresponding subjective impressions.

These irregularities led to regression formulas of considerably less accuracy [about 60 percent of variance accounted for by eqs. (28) and (29)] than the 70 to 90 percent obtained in other studies.[4,13] Our work lends support to the value of current efforts to find more robust objective measures.[17–19]

VII. ACKNOWLEDGMENT

We are grateful to Barbara McDermott for advice and assistance with the experimental design and the data processing.

REFERENCES

1. R. E. Crochiere, D. J. Goodman, L. R. Rabiner, and M. R. Sambur, "Tandem Connections of Wideband and Narrowband Speech Communications Systems: Part 1—Narrowband-to-Wideband Link," B.S.T.J., *56*, No. 9 (November 1977), pp. 1701–1722.
2. L. R. Rabiner, M. R. Sambur, R. E. Crochiere, and D. J. Goodman, "Tandem Connections of Wideband and Narrowband Speech Communications Systems: Part 2—Wideband-to-Narrowband Link," B.S.T.J., *56*, No. 9 (November 1977), pp. 1723–1741.
3. P. Noll, "Adaptive Quantizing in Speech Coding Systems," Int. Zurich Seminar, March 1974.
4. B. J. McDermott, C. Scagliola, and D. J. Goodman, "Perceptual and Objective Evaluation of Speech Processed by Adaptive Differential PCM," B.S.T.J., *57*, No. 5 (May–June, 1978), pp. 1597–1618.
5. R. E. Crochiere and L. R. Rabiner, "Optimum FIR Digital Filter Implementations for Decimation, Interpolation, and Narrowband Filtering," IEEE Trans. Acoust., Speech, Sig. Proc., *ASSP-23* (October 1975), pp. 444–456.
6. L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," IEEE Trans. Acoust., Speech, and Sig. Proc., *ASSP-25*, No. 1 (February 1977), pp. 24–33.
7. J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-Time Hardware Pitch Detector," IEEE Trans. Acoust., Speech, and Sig. Proc., *ASSP-24*, No. 1 (February 1976), pp. 2–8.
8. M. R. Sambur, "An Efficient Linear Prediction Vocoder," B.S.T.J., *54*, No. 10 (December 1975), pp. 1693–1723.
9. R. A. McDonald, "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems," B.S.T.J., *45*, No. 7 (September 1966), pp. 1123–1151.
10. R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Sub-bands," B.S.T.J., *55*, No. 8 (October 1976), pp. 1069–1086.
11. R. E. Crochiere, "On the Design of Sub-band Coders for Low-Bit-Rate Speech Communication," B.S.T.J., *56*, No. 5 (May–June 1977), pp. 747–770.
12. R. E. Crochiere, "An Analysis of 16 kb/s Sub-band Coder Performance: Dynamic

Range, Tandem Connections, and Channel Errors," B.S.T.J., *57,* No. 8 (October 1978), pp. 2927–2952.
13. D. J. Goodman, B. J. McDermott, and L. H. Nakatani, "Subjective Evaluation of PCM Coded Speech," B.S.T.J., *55,* No. 8 (October 1976), pp. 1087–1109.
14. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-23,* No. 1 (February 1975), pp. 67–72.
15. A. H. Gray, Jr. and J. D. Markel, "Distance Measures for Speech Processing," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-24* (October 1976), pp. 380–391.
16. N. H. Nie, C. H. Hull, J. H. Jenkins, K. Steinbrenner, and D. H. Bent, *Statistical Package for the Social Sciences,* second ed., New York: McGraw Hill, 1975, pp. 426–428.
17. C. Scagliola, "Evaluation of Adaptive Speech Coders Under Noisy Channel Conditions," Int'l Conf. on Commun., June 1978.
18. R. E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A Study of Objective Measures for Speech Waveform Quantization," Int'l Zurich Seminar, 1978.
19. J. M. Tribolet, P. Noll, B. J. McDermott, and R. E. Crochiere, "A Study of Complexity and Quality of Speech Waveform Coders," Proceedings of the 1978 IEEE International Conf. on Acoustics, Speech and Signal Proc., Tulsa, Okla., April 1978.